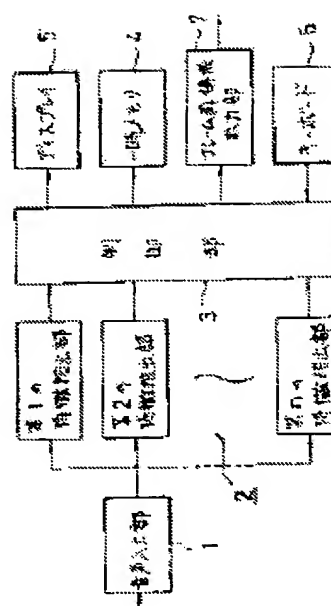


# SPEECH ELEMENT FILE PRODUCING DEVICE

**Patent number:** JP1244499  
**Publication date:** 1989-09-28  
**Inventor:** SAITO SHIGETOSHI  
**Applicant:** TOSHIBA CORP  
**Classification:**  
**- international:** G10L5/04; G06F3/16  
**- european:**  
**Application number:** JP19880069787 19880325  
**Priority number(s):**

## Abstract of JP1244499

**PURPOSE:** To produce a high-quality voice file by displaying the information of feature parameters found by acoustically analyzing input voices in graphs and registering an arbitrary frame to the displayed data after expansion or contraction.  
**CONSTITUTION:** Input voices fetched through a voice input section 1 are acoustically analyzed at a feature extracting section 2 and their acoustic feature is found. A control section 3 stores the information sequence of the feature parameters in a temporary memory 4 and, at the same time, graphically displays the information sequence by using a display 5. An operator checks the information from the display. When control information for producing a speech element file is inputted thereafter, the control section 3 gives the instruction information to a frame reconstitution output section 7. The output section 7 reads out the information sequence of the feature parameters of the input voices from the memory 4 in accordance with the instruction information and performs expansion and contraction processes on arbitrary frames so as to reconstitute the arrangement of the frames. The sequence of the feature parameters to which the processes are performed is extracted as speech element parameters.



Data supplied from the esp@cenet database - Worldwide

BEST AVAILABLE COPY

⑨ 日本国特許庁(JP)

⑪ 特許出願公開

⑫ 公開特許公報(A) 平1-244499

⑤ Int. Cl.<sup>4</sup>

識別記号

庁内整理番号

⑬ 公開 平成1年(1989)9月28日

G 10 L 5/04  
G 06 F 3/16

3 3 0

E-8622-5D  
K-7341-5B

審査請求 未請求 請求項の数 1 (全5頁)

⑭ 発明の名称 音声素片ファイル作成装置

⑯ 特 願 昭63-69787

⑰ 出 願 昭63(1988)3月25日

⑱ 発 明 者 齊 藤 成 利 神奈川県川崎市幸区柳町70番地 株式会社東芝柳町工場内

⑲ 出 願 人 株 式 会 社 東 芝 神奈川県川崎市幸区堀川町72番地

⑳ 代 理 人 弁 理 士 鈴 江 武 彦 外 2 名

明 細 書

1. 発明の名称

音声素片ファイル作成装置

2. 特許請求の範囲

音声合成に用いられる音韻パラメータ系列を、文字列を分析して求められる音韻系列に応じて生成する為の音声素片ファイルを作成する音声素片ファイル作成装置において、

入力音声を音響分析してその特徴パラメータを求める音響分析部と、この音響分析データをグラフィック表示する表示部と、このグラフィック表示された音響分析データに対する任意のフレームの拡張・圧縮を指示する為の指示入力部と、指示入力された情報に従って前記音響分析された特徴パラメータの任意のフレームを拡張・圧縮して抽出し、音声素片ファイルとして登録する手段とを具備したことを特徴とする音声素片ファイル作成装置。

3. 発明の詳細な説明

[発明の目的]

(産業上の利用分野)

本発明は音声合成に用いられる音韻パラメータ系列を音韻系列に応じて生成する為の音声素片ファイルを効果的に作成し得る音声素片ファイル作成装置に関する。

(従来の技術)

マンマシン・インターフェースを実現する重要な技術の1つに音声合成技術がある。この音声合成は、例えば入力文字コード列を解析してその音韻パラメータ系列と韻律パラメータ系列とを求め、これ等のパラメータ系列に従って声道特性を近似するディジタルフィルタを所定の規則に基づいて構築し、前記入力文字コード列に対応した合成音声を生成するものである。

このような規則合成法による音声合成に必要な上記音韻パラメータ系列を作成する為に音声素片ファイルが用いられる。

ところで結合の音声素片の単位としては、一般

に音素片、子音・母音（C V）素片、母音・子音・母音（V C V）素片、子音・母音・子音（C V C）素片等が用いられる。しかし音声合成に用いられる上述した音素片をファイルとして予め登録しておく為、従来では孤立発声されるC V音節（C V, V）や単語、短文章の中から、求めようとする音素片を切出し、その特徴パラメータ（ケプストラム、メルケプストラム、L S P, P A R C O R等）を音響分析によって求め、これを音素片ファイルとして登録している。

然し乍ら、このようにして求められる音素片ファイルにあっては、発声入力された音声の母音部のゆらぎや、子音長が短すぎることに起因して、音声合成に用いられる音響パラメータ系列の生成に供される音素片として十分な特徴パラメータが得られない等の不具合が生じた。この結果、明瞭度が高く、母音部の安定した合成音声を得ることが難しくなると云う問題があった。

（発明が解決しようとする課題）

このように従来にあっては孤立発声された音

— 3 —

れる情報に従って、前記音響分析された特徴パラメータにおける任意のフレームを拡張・圧縮して抽出し、これを音素片ファイルとして登録するようにしたことを特徴とするものである。

（作用）

本発明によればグラフィック表示される特徴パラメータの情報（スペクトル、スペクトル変化量、有声／無声情報、音声パワー、音声波形等）を視察して任意のフレームに対する拡張・圧縮を指示すれば、入力音声を音響分析して求められた特徴パラメータの系列における任意のフレームがその指示情報に従って拡張・圧縮されて抽出され、この処理が施された特徴パラメータの情報が音素片ファイルとして登録される。

この結果、発声入力された音声における母音部のゆらぎを吸収し、また子音長の不足を補った特徴パラメータの系列を生成し、これを音素片ファイルとして登録することが可能となる。そしてこの音素片ファイルを用いることで、明瞭度が高く母音部の安定した聞き取り易い高品質な合成音声

— 5 —

声等から音素片を切出し、その音素片の特徴パラメータを求めて音素片ファイルに登録しているだけなので、入力音声の母音部のゆらぎや、子音長が短いこと等に起因して、明瞭度の高い安定した合成音声を生成する為の品質の高い音素片ファイルを求めることが困難であった。

本発明はこのような事情を考慮してなされたもので、その目的とするところは、明瞭度の高い安定した合成音声の生成の基本となる品質の高い音素片ファイルを簡易に生成することのできる音素片ファイル作成装置を提供することにある。

〔発明の構成〕

（問題点を解決するための手段）

本発明に係る音素片ファイル作成装置は、音響分析部にて入力音声を音響分析して求められる特徴パラメータの情報（スペクトル、スペクトル変化量、有声／無声情報、音声パワー、音声波形等）を表示部にてグラフィック表示し、このグラフィック表示された音響分析データを視察し、その視察結果に応じてオペレータから指示入力さ

— 4 —

を生成することが可能となる。

（実施例）

以下、図面を参照して本発明の一実施例につき説明する。

第1図は実施例装置における音素片ファイル生成処理手続きの流れを示す図であり、第2図は実施例装置の要部概略構成図である。

第2図において1は発声入力される音声を取込む音声入力部であり、マイクロフォンや増幅器等を備えて構成される。この音声入力部1を介して取込まれた入力音声は特徴抽出部2に導かれて音響分析され、その音響的な特徴が所定のフレーム単位で求められる。この音響分析部2は、例えば複数の分析手段からなり、入力音声のスペクトル、スペクトル変化量、有声／無声情報、音声パワー、音声波形等の音響的特徴をそれぞれ求める。

制御部3は音響分析された入力音声の特徴パラメータの情報系列を一時メモリ4に格納すると共に、その特徴パラメータの情報系列をディスプレイ5を用いて高速グラフィック表示する。この特

— 6 —

微パラメータの情報系列の高速グラフィック表示によって入力音声を持つ情報（音響的性質）の視察がオペレータにより行われ、その入力音声からの音声素片ファイルの作成が制御指示される。

しかしてキーボード6から音声素片ファイル作成の為の制御情報が指示入力されると、制御部3はその指示情報をフレーム再構成出力部7に与え、該フレーム再構成出力部7を起動する。するとフレーム再構成出力部7は上記指示入力された制御情報に従って前記一時メモリ4から入力音声の特徴パラメータの情報系列を読出し、任意のフレームに対して拡張・圧縮の処理を施し、上記特徴パラメータ系列のフレームの並びを再構成する。この任意のフレームの拡張・圧縮処理が施された特徴パラメータの系列が音声素片ファイルとして抽出される。そしてこの情報がファイルに登録されて音声合成処理における音韻パラメータ系列の生成に用いられる。

即ち、本装置では第1図に示すように音声素片ファイルの生成に用いる為の音声を、孤立発声単語

- 7 -

に基づいて任意のフレームに対する拡張・圧縮を指示し、音声素片の切出し作成を制御指示する（ステップd）。

例えば第3図に示すグラフィック表示データからは、そのスペクトラム変化量から、/ma/のCVわたり位置としては9フレーム目が適當であること、また母音部分として10～22フレーム目の12フレーム分を切出すと、20～22フレーム目の音声パワー値が減少していることから音声素片としては不適切であることがわかる。

そこでこのような場合には9フレーム目をCVわたり位置として、また11、12フレームを4回繰返して母音部分とすることをキーボード6から制御情報として指示入力する。この制御情報を受けて指示されたフレームについての拡張・圧縮処理が施される（ステップe）。つまり第6図に示すように11、12フレームの特徴情報を4回繰返したフレーム構成の特徴パラメータ系列が変換生成される。この結果、例えば第4図に示すようにフレームの再構成がなされ、その母音部分の

- 9 -

等として音声入力し（ステップa）、その入力音声を音響分析する（ステップb）。この音響分析処理によって前述したような種々の特徴パラメータの時系列が求められる。

しかる後、音響分析により求められた特徴情報を、例えば第3図に示すように高速グラフィック表示する（ステップc）。尚、ここでは孤立発声させた男性音声/ma/の分析データとして、スペクトル変化量、音声パワー、有声/無声情報、スペクトル、音声波形を例示しているが、その他の分析情報をグラフィック表示することも可能である。また第3図に示す分析情報の全てを表示する必要がないことも勿論のことである。

またここでの音声パワー、有声/無声情報、音声波形は入力音声を直接分析することにより求められるが、スペクトルやスペクトル変化量は、例えば音声パワーのケプストラム・パラメータから計算する等して求められる。

このようにしてグラフィック表示される入力音声の特徴情報に対して、オペレータは種々の経験

- 8 -

安定化が図られた特徴パラメータの系列が求められ、この特徴パラメータ系列が上記/ma/の音声素片として抽出されて音声素片ファイルに登録される（ステップf）。

尚、任意のフレームの拡張に関しては上述した2フレームを一単位とする繰返しのみならず、第7図に示すように1フレームを一単位とする繰返しを指定することも可能である。例えば第3図に示す/ma/の特徴パラメータの系列に対して、/m/の中間フレームである6フレーム目の4フレームに互る拡張を指示した場合には、第7図に示すようなフレームの拡張処理が行われ、子音長の短さを補った第5図に示すような特徴パラメータ系列が求められて音声素片ファイルに登録される。

このようなフレームの拡張のみならず、長すぎる母音部分を圧縮するような場合には、第8図に示すようなフレームの間引きによる圧縮処理を施すようにすれば良い。

以上説明したように本装置によれば、グラフィ

- 10 -

ック表示された音声の特徴情報に従って任意のフレームの拡張・圧縮を指示し、この指示情報に従う入力音声の特徴パラメータの系列に対する任意のフレームの拡張・圧縮処理によって母音部分のゆらぎを吸収し、或いは子音長の短さを補った音声素片を効果的に作成することができる。この結果、様々な環境で発声入力された音声から、明瞭度が高く、母音部分のゆらぎの少ない高品質な合成音声を生成する為の音声素片ファイルを得ることが可能となる。

尚、本発明は上述した実施例に限定されるものではない。ここではC V音節を孤立発声させた音声からの音声素片ファイルの作成につき説明したが、孤立発声単語や短文章中から音声素片ファイルを切抜き抽出して作成することも可能である。また音声素片の単位も音素片(C, V)やC V素片, V C V素片, C V C素片等として実現できる。また音声の特徴パラメータとしても上述した例に限定されず、また種々の特徴パラメータを適宜組合わせて用いることができる。その他、本発明は

- 1 1 -

その要旨を逸脱しない範囲で種々変形して実施することができる。

#### 〔発明の効果〕

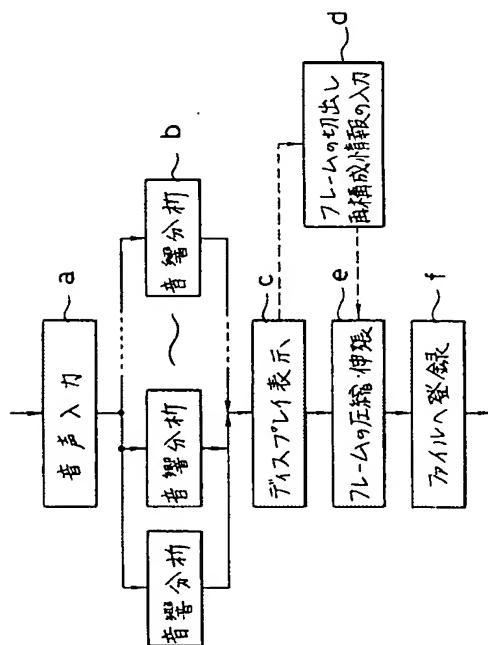
以上説明したように本発明によれば、種々の環境下で入力された音声から母音部分のゆらぎや子音長の短過ぎに起因する不具合のない、明瞭度の高い高品質な合成音声を得ることのできる音声素片ファイルを簡易に、且つ効果的に生成することができる等の実用上多大なる効果を奏し得る。

#### 4. 図面の簡単な説明

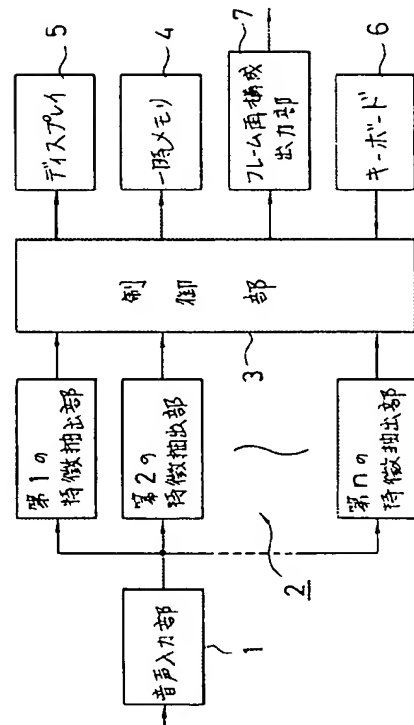
図は本発明に係る音声素片ファイル作製装置の一実施例を示すもので、第1図は特徴的な処理手続きの流れを示す図、第2図は実施例装置の要部概略構成図、第3図乃至第5図はグラフィック表示される音声の分析データとその修正例を示す図、第6図乃至第8図はフレームの拡張・圧縮の概念を示す図である。

1…音声入力部、2…特徴抽出部、3…制御部、4…一時メモリ、5…ディスプレイ、6…キーボード、7…フレーム再構成出力部。

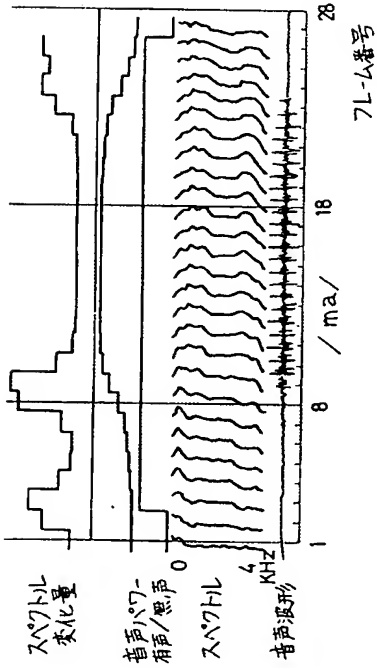
- 1 2 -



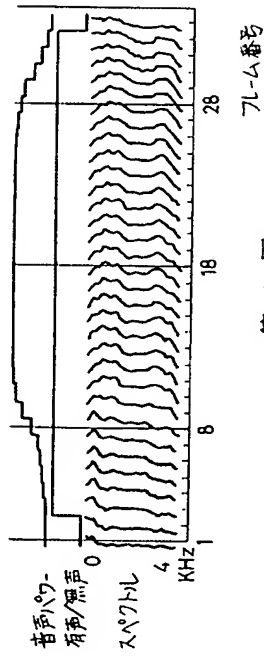
第1図



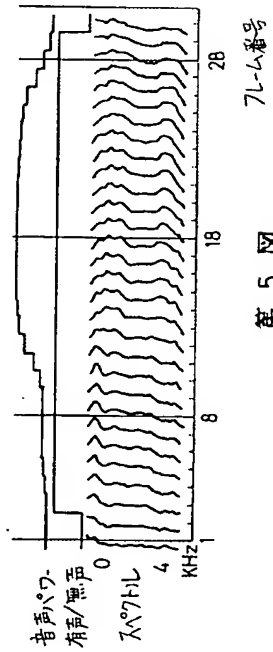
第2図



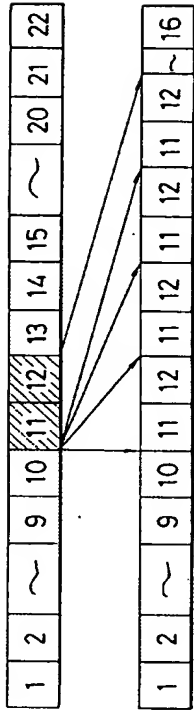
第 3 図



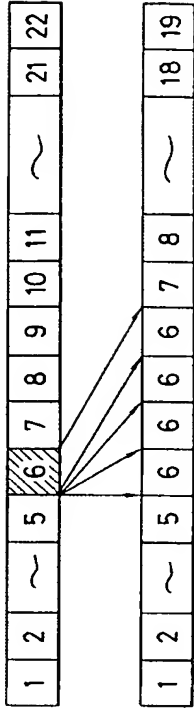
第 4 図



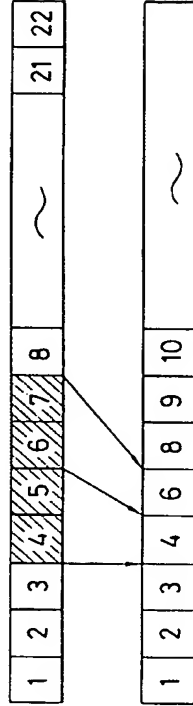
第 5 図



第 6 図



第 7 図



第 8 図